

**ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED  
UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP  
KENAIKAN HARGA ROKOK PADA MEDIA SOSIAL  
TWITTER**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada Jurusan  
Informatika Fakultas Komunikasi dan Informatika**

**Oleh:**

**IIN KUSUMAWATI**

**L 200 130 008**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA  
2017**

## **HALAMAN PERSETUJUAN**

# **ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP KENAIKAN HARGA ROKOK PADA MEDIA SOSIAL TWITTER**

## **PUBLIKASI ILMIAH**

oleh:

**IIN KUSUMAWATI**

**L 200 130 008**

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



**Endang Wahyu Pamungkas, S.Kom., M.Kom.**

**NIK. 100.1704**

## HALAMAN PENGESAHAN

# ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP KENAIKAN HARGA ROKOK PADA MEDIA SOSIAL TWITTER

OLEH

IIN KUSUMAWATI

L 200 130 008

Telah dipertahankan di depan Dewan Penguji Fakultas  
Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Pada hari Selasa, 24 Januari 2017 dan  
dinyatakan telah memenuhi syarat

Dewan Penguji:

1. Endang Wahyu Pamungkas, S.Kom., M.Kom.  
(Ketua Dewan Penguji)
2. Helman Muhammad, S.T., M.T..  
(Anggota I Dewan Penguji)
3. Aris Rakhmadi, S.T., M.Eng..  
(Anggota II Dewan Penguji)

  
(.....)  
  
(.....)  
  
(.....)

Dekan,  
Fakultas Komunikasi dan Informatika



Husni Thamrin, S.T., M.T., Ph.D.

NIK. 706

Ketua Program Studi Informatika



Dr. Heru Supriyono, S.T., M.Sc.

NIK. 970

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam Naskah Publikasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

**Surakarta, 24 Januari 2017**

Penulis



**IIN KUSUMAWATI**

**L 200 130 008**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

---

**SURAT KETERANGAN LULUS PLAGIASI**

**012/A.3-II.3/INF-FKI/I/2017**

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : IIN KUSUMAWATI  
NIM : L200130008  
Judul : ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED  
UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP  
KENAIKAN HARGA ROKOK PADA MEDIA SOSIAL TWITTER  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 02 Februari 2017

Biro Skripsi Informatika

**Ihsan Cahyo Utomo, S.Kom., M.Kom.**



UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

wisuda 2017 wisuda maret - DUE 17-Jan-2017 Roadmap Paper 15 of 38

Originality GradeMark PeerMark

ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED UNTUK MELIHAT  
BY IIN KUSUMAWATI

turnitin 14% SIMILAR OUT OF 0

Match Overview

1	Submitted to Universit...	6%
2	airccj.org	1%
3	dergipark.ulakbim.gov.tr	1%
4	airccse.org	1%
5	eprints.ums.ac.id	1%
6	Submitted to Universit...	1%
7	labti.ukdw.ac.id	1%
8	openlibrary.heatonrese...	1%

**ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP KENAIKAN HARGA ROKOK PADA MEDIA SOSIAL TWITTER**

**Abstrak**

Perkembangan teknologi memberi dampak pada kemudahan mengakses dan membagi informasi, terutama media sosial twitter. Pesan yang dikirim pada twitter dapat berupa berita terkini, gagasan, opini maupun curahan hati. Menurut situs *alexa*, *twitter* berada pada peringkat kedelapan situs global. Dan menurut situs *eBizMBA*, pengguna *twitter* pada periode September 2016 mencapai jumlah 310 juta. Informasi yang tertuang pada media sosial sebagian besar berupa opini. Dari banyaknya opini, dapat dimanfaatkan pihak-pihak tertentu untuk mengevaluasi keadaan kedepannya. Keadaan ini dapat berupa keadaan negatif maupun positif. Maka dari itu, diperlukan adanya teknik untuk dapat menggolongkan opini sesuai sentimen positif atau sentimen negatif. Teknik yang digunakan yaitu *opinion mining* dan *sentiment analysis*. Salah satu metode dari *sentiment analysis* yaitu *lexicon based*. *Lexicon based* didasarkan pada orientasi kontekstual sentimen pada jumlah orientasi sentimen pada setiap kata atau kalimat. Oleh karena itu, penelitian ini mencoba menganalisis persepsi masyarakat kedalam kelas sentimen menggunakan metode *lexicon based* dengan SentiWordNet *Dataset* yang digunakan adalah *tweets* mengenai kenaikan harga rokok dalam bahasa indonesia berjumlah 350 buah. Data diklasifikasikan sesuai SentiWordNet pada tiap-tiap kata dalam kalimat. Untuk kata yang memiliki lebih dari satu arti maka *synset* dipilih berdasarkan metode First Sense dari SentiWordNet yang muncul paling populer. Hasil dari penelitian ini adalah perolehan nilai *accuracy* tertinggi sebesar 81% untuk tipe opini netral. Sedangkan nilai *precision* dan *recall*, tipe opini negatif mempunyai nilai lebih tinggi dibandingkan tipe opini lain dengan nilai *precision* yaitu 53% dan nilai *recall* sebesar 57%.

# ANALISA SENTIMEN MENGGUNAKAN LEXICON BASED UNTUK MELIHAT PERSEPSI MASYARAKAT TERHADAP KENAIKAN HARGA ROKOK PADA MEDIA SOSIAL TWITTER

## Abstrak

Perkembangan teknologi memberi dampak pada kemudahan mengakses dan membagi informasi, terutama media sosial twitter. Pesan yang dikirim pada twitter dapat berupa berita terkini, gagasan, opini maupun curahan hati. Menurut situs *alexa*, *twitter* berada pada peringkat kedelapan situs global. Dan menurut situs *eBizMBA*, pengguna *twitter* pada periode September 2016 mencapai jumlah 310 juta. Informasi yang tertuang pada media sosial sebagian besar berupa opini. Dari banyaknya opini, dapat dimanfaatkan pihak-pihak tertentu untuk mengevaluasi keadaan kedepannya. Keadaan ini dapat berupa keadaan negatif maupun positif. Maka dari itu, diperlukan adanya teknik untuk dapat menggolongkan opini sesuai sentimen positif atau sentimen negatif. Teknik yang digunakan yaitu *opinion mining* dan *sentiment analysis*. Salah satu metode dari *sentiment analysis* yaitu *lexicon based*. *Lexicon based* didasarkan pada orientasi kontekstual sentimen pada jumlah orientasi sentimen pada setiap kata atau kalimat. Oleh karena itu, penelitian ini mencoba menganalisis persepsi masyarakat kedalam kelas sentimen menggunakan metode *lexicon based* dengan SentiWordNet. *Dataset* yang digunakan adalah *tweets* mengenai kenaikan harga rokok dalam bahasa indonesia berjumlah 350 buah. Data diklasifikasikan sesuai SentiWordNet pada tiap-tiap kata dalam kalimat. Untuk kata yang memiliki lebih dari satu arti maka synset dipilih berdasarkan metode First Sense dari SentiWordNet yang muncul paling populer. Hasil dari penelitian ini adalah perolehan nilai *accuracy* tertinggi sebesar 81% untuk tipe opini netral. Sedangkan nilai *precision* dan *recall*, tipe opini negatif mempunyai nilai lebih tinggi dibandingkan tipe opini lain dengan nilai *precision* yaitu 53% dan nilai *recall* sebesar 57%.

**Kata kunci:** lexicon-based, opinion mining, sentiment analysis, sentiwordnet, tweets

## Abstract

Technological developments give impact on ease of access and share information, especially social media twitter. A message posted on twitter can be the breaking news, ideas, opinions or outpouring of the heart. According to the alexa website, twitter ranks eighth globally. According to the site, there were almost 310 million users on eBizMBA during September 2016. The most information in the social media is about opinion. From a large number of opinions, can be utilized to evaluate the State of the future of certain parties. This State can be a negative and a positive State of affairs. Thus, the existence of the necessary techniques to be able to classify appropriate opinion sentiment positive or negative sentiment. Techniques employed i.e. opinion mining and sentiment analysis. One method of sentiment analysis i.e. lexicon-based. Lexicon-based based on the contextual orientation of sentiment on the number orientation of any word or sentence. Therefore, this study tries to analyze the perceptions of the community into the classroom using sentiment lexicon-based method with SentiWordNet. The dataset used in this research are the tweets about the price increase smoking in Indonesia totaled 350 pieces. The data are classified in accordance SentiWordNet in each word in the sentence. For words that have more than one meaning then selected based on the method of First synset Sense of SentiWordNet that appears the most popular. Results from this study is the acquisition value of the highest accuracy of 81% for the type of neutral opinion. While the value of precision and recall, the type of negative opinion has more value than other types of opinion with the value at 53% precision and 57% recall value.

**Keywords:** lexicon-based, opinion mining, sentiment analysis, sentiwordnet, tweets

## 1. PENDAHULUAN

Dewasa ini, perkembangan teknologi berkembang secara drastis dan terus berevolusi sampai sekarang. Hal tersebut memberi dampak pada kemudahan mengakses dan membagi informasi. Terutama pada penggunaan internet untuk akses ke situs media sosial. Akses ke media sosial ini banyak merubah gaya komunikasi manusia pada sesamanya. Adanya *twitter* merupakan salah satu media sosial yang memungkinkan penggunanya mengirim pesan baik berupa berita terkini, gagasan, opini ataupun curahan hati. Pesan ini terkenal dengan sebutan *tweet*. *Tweet* merupakan pesan pendek yang hanya berjumlah 140 karakter. Menurut situs *alexa*, *twitter* berada pada peringkat kedelapan situs global. Dan menurut situs *eBizMBA*, pengguna *twitter* pada periode September 2016 mencapai jumlah 310 juta (Purba, Hidayati & Gozali, 2014). Banyaknya jumlah pengguna *twitter* kerap dimanfaatkan instansi atau penjual untuk meninjau opini pengguna mengenai isu ataupun barang yang dijual. Apakah mendapat evaluasi yang baik atau malah menjadi buruk.

Informasi yang tertuang pada media sosial sebagian besar berupa opini. Pada situs *Webster's New Collegiate Dictionary*, opini merupakan sebuah pemikiran mengenai penilaian terhadap suatu permasalahan (Purba, Hidayati & Gozali, 2014). Siapapun bisa memiliki ide atau gagasan yang sama terhadap suatu fakta, namun anggapan atau opini terhadap suatu hal bisa menjadi berbeda-beda. Dari banyaknya opini, dapat diketahui bahwa tidak hanya mengandung keadaan yang positif tetapi juga mengandung keadaan negatif. Terlebih pada penelitian (Purnamasari, Adji & Buntoro, 2014) mengkategorikan menjadi 7 parameter yaitu sangat positif, positif, agak positif, netral, agak negatif, negatif, sangat negatif. Sehingga diperlukan adanya teknik untuk dapat menggolongkan opini sesuai sentimen yang cocok. *Opinion mining* dan *sentiment analysis* (Pang & Lee, 2008) merupakan teknik yang kerap dipakai untuk menentukan opini, sentimen, dan subjektivitas dalam teks. Proses *mining* sendiri merupakan proses pengolahan data dengan memfungsikan algoritma dan metode yang telah ditentukan sebelumnya (Listiana, 2015).

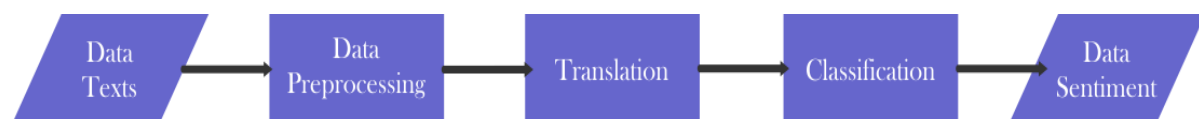
*Sentiment Analysis* merupakan sebuah penggambaran polaritas pada suatu teks atau kata (Esuli & Sebastiani, 2006). Metode ini mengekstraksi pendapat dari seseorang mengenai sesuatu hal seperti produk atau kandidat politik. Salah satu penelitian dengan menggunakan metode *sentiment analysis* pada media sosial *twitter* yaitu "Sentiment Analysis on Twitter" (Kumar & Sebastian, 2012). Penelitian dilakukan untuk membuat dan menggunakan teori dan teknologi dengan metode berbasis corpus. Selain itu, *sentiment analysis* dengan data *twitter* juga dapat memprediksi isu sosial kota Bandung (Purba, Hidayati & Gozali, 2014). *Sentiment Analysis* dapat dilakukan dengan salah satu cara yaitu *lexicon based*. Pendekatan ini didasarkan



pada orientasi kontekstual sentimen pada jumlah orientasi sentimen pada setiap kata atau kalimat (Turney 2002). *Lexicon based* umumnya menggunakan kamus untuk mendukung klasifikasi sentimen yaitu SentiWordNet. SentiWordNet merupakan hasil anotasi otomatis dari semua synset pada WordNet (Baccianella, Esuli, dan Sebastiani 2010). WordNet disebut juga sebagai database leksikal untuk bahasa Inggris. Pada WordNet, kata dikelompokkan ke dalam set sinonim yang disebut synset.

Pada tugas akhir ini, penelitian akan berpusat pada perbandingan hasil *sentiment analysis* mengenai persepsi masyarakat terhadap kenaikan harga rokok yang sudah diklasifikasi secara manual dengan klasifikasi yang dihasilkan oleh sistem. Penelitian dibutuhkan karena adanya wacana kebijakan baru dari pemerintah. Berdasarkan masalah ini, penelitian dilakukan menggunakan metode *lexicon based* dengan SentiWordNet. Data yang diperoleh dalam penelitian ini merupakan data dari twitter dalam bahasa Indonesia. Oleh karena itu, dilakukan translasi dahulu untuk mengubah Bahasa Indonesia menjadi Bahasa Inggris karena SentiWordNet hanya memberikan fasilitas Bahasa Inggris.

## 2. METODOLOGI



Penelitian ini dilakukan dalam beberapa tahap. Pada tahap pengumpulan data, data berupa kalimat-kalimat yang diperoleh dari *tweets* berbahasa Indonesia mengenai hasil persepsi masyarakat. Setelah data terkumpul kemudian dilakukan *preprocessing* agar data dapat diolah menggunakan *sentiment analysis*. Proses translasi ke Bahasa Inggris dilakukan pada kalimat yang sudah melewati tahap *preprocessing*. Selanjutnya mengklasifikasi kalimat dengan analisis sentiment sehingga output berupa nilai sentiment. Berikut penjelasan tahap-tahap dari diagram.

### 2.1 Data Preprocessing

Tujuan dari tahap *preprocessing* adalah menormalkan teks menjadi bentuk yang tepat. Sehingga data yang sebelumnya masih mentah dapat diproses menggunakan *sentiment analysis* menjadi data yang berkualitas. Pada tahap ini praproses dilakukan secara manual. Berikut langkah-langkahnya :

#### 2.1.1 Deteksi hastag dan mention

Data diambil dari *tweets* pada media sosial <https://twitter.com/> dengan menggunakan *query* pencarian mengenai kenaikan harga rokok. Tweets yang diambil mengandung hastag

#rokokmahal, #POLEMIKHargaRokok, #ILCrokak50ribu, #kenaikanhargarokok, #rokok dan mention kepada user @jokowi, @ilc\_tvOnenews. Data yang diambil hanya *tweets* berbahasa Indonesia. Selain itu, tidak mengambil *tweets* yang muncul dari hastag pada situs berita karena bukan merupakan pandangan atau opini dari masyarakat.

### 2.1.2 Penghapusan karakter

Kalimat-kalimat yang diperoleh dari *twitter* biasanya masih terdapat kesalahan secara acak atau varian dalam variabel. Untuk itu, kesalahan atau varian ini harus dihapus. Kata yang dihapus antara lain karakter HTML, hashtag(#), username(@), url(http://namasitus.com), dan email (nama@domain.com)

### 2.1.3 Penggantian kata

Penggantian kata atau *replacement* dilakukan untuk membetulkan kata pada kalimat yang tidak padu dan belum efektif. Banyak pengguna yang menggunakan singkatan-singkatan dan ragam bahasa yang tidak resmi seperti bahasa slang atau bahasa gaul. Singkatan-singkatan ini dirubah menjadi frase atau nama sesuai bentuk aslinya. Kemudian pada penggunaan bahasa slang atau bahasa gaul, perubahan bentuk pesan dengan maksud penyembunyian atau kejenakaan diganti menjadi bahasa baku tanpa mengubah isinya. Selain itu, juga dilakukan pengubahan setiap karakter huruf dengan huruf kecil. Aturan dalam proses penggantian kata dapat dilihat pada Tabel 1.

Tabel 1. Penggantian kata pada singkatan dan bahasa slang atau gaul

Singkatan dan Bahasa Gaul	Normal
Kosakata khas: gue, elu	Kata: saya, kamu
Kata berulang: hati2	Kata berulang: hati-hati
Penghubung: yg	Penghubung: yang
Singkatan: Rp	rupiah
Singkatan: dst	dan seterusnya
Akhiran -e	Akhiran -a
Akhiran -ny	Akhiran -nya
Akhiran -nk	Akhiran -ng
Preposisi d-	Preposisi di-
Preposisi k-	Preposisi ke-

### 2.1.4 Memperpendek kata yang berlebihan

Kata-kata yang memiliki huruf yang sama lebih dari dua kali direduksi menjadi kata berulang yang terjadi hanya sekali. Seperti mengurangi kelebihan huruf pada kata “JANGAAAAAN” menjadi “JANGAN”, “asaalll” menjadi “asal”.

Hasil dari praproses akan dilanjutkan pada tahap berikutnya yaitu Translation.

### 2.2 Translation

Data hasil praproses kemudian dilakukan proses penerjemahan ke bahasa Inggris. Tahap ini menggunakan layanan *translator* di web. Salah satu *translator* yang digunakan untuk proses penerjemahan ini yaitu Bing Translator. Web dari microsoft ini memiliki tingkat akurasi terjemahan bahasa inggris yang cukup bagus. Meskipun begitu, pengeditan secara manual masih perlu dilakukan.

Hasil dari proses terjemahan akan digunakan untuk tahap selanjutnya yaitu *Classification*.

### 2.3 Classification

Dalam proses klasifikasi ini, penentuan data dilakukan pada tiap kata dalam kalimat menggunakan *lexicon based* dengan SentiWordNet. Untuk kata yang memiliki lebih dari satu arti maka synset akan dipilih berdasarkan metode First Sense dari SentiWordNet yang muncul paling atas atau populer. Kemudian, kata yang terklasifikasi sesuai SentiWordNet dilakukan pencarian nilai sentimen dalam satu kalimat dengan rumus.

$$S_{positive} = \sum_{i \in t}^n positive\ score_i \quad (1)$$

$$S_{negative} = \sum_{i \in t}^n negative\ score_i \quad (2)$$

Dimana ( $S_{positive}$ ) adalah bobot dari kalimat yang didapatkan melalui penjumlahan n skor polaritas kata opini positif dan ( $S_{negative}$ ) adalah bobot dari kalimat yang didapatkan melalui penjumlahan n skor polaritas kata opini negatif. Bobot pada tiap kalimat ini yang akan digunakan sebagai acuan untuk melakukan proses perbandingan. Sehingga dalam satu kalimat akan diketahui total jumlah nilai positif ( $S_{positive}$ ) dan juga nilai negatif ( $S_{negative}$ ) dari tiap-tiap kata penyusunnya. Dari persamaan nilai sentimen dalam satu kalimat maka diperoleh persamaan 3 untuk menentukan orientasi sentimen dengan perbandingan jumlah nilai positif, negatif dan netral.

$$Sentence_{sentiment} = \begin{cases} positive & \text{if } S_{positive} > S_{negative} \\ neutral & \text{if } S_{positive} = S_{negative} \\ negative & \text{if } S_{positive} < S_{negative} \end{cases} \quad (3)$$

Jika total jumlah nilai positif lebih besar dari jumlah nilai negatif maka kalimat akan berorientasi positif. Jika total jumlah nilai positif sama dengan jumlah nilai negatif maka kalimat akan berorientasi objektif atau netral. Namun jika total jumlah nilai positif kurang dari jumlah nilai negatif maka kalimat akan berorientasi negatif (Pamungkas & Putri, 2016).

### 3. HASIL DAN PEMBAHASAN

Ujicoba dilakukan dengan dataset yang terdiri dari *realtime tweet* berupa opini berbahasa Indonesia berjumlah 350 buah. Data diambil secara manual dan disimpan dalam database sebagai *gold standart* data. Pengambilan data dilakukan menggunakan *query* pencarian seperti pada tabel dibawah.

Tabel 2. Query pencarian

Hastagh dan Mention	Jumlah
#rokok	240
#rokokmahal	30
#POLEMIKHargaRokok	10
#ILCrokok50ribu	15
#kenaikanhargarokok	10
@jokowi	25
@ilc_tvOnenews	20

Dalam *gold standart* data, dilakukan klasifikasi sentimen untuk tiap kalimat dengan pemberian label positif, negatif, dan netral. Ujicoba menggunakan bahasa pemrograman Java untuk penerapan metode.

Tabel 3. Hasil pengujian dengan metode *lexicon based*

Opinion Type	Precision	Recall	Accuracy
Positive	51%	56%	58%
Negative	53%	57%	60%
Neutral	10%	4%	81%

Tabel diatas adalah hasil ujicoba pada data untuk tiap tipe opini. *Precision*, *recall* dan *accuracy* merupakan parameter yang digunakan untuk mengetahui seberapa baik kualitas dari pengujian. *Precision* adalah tingkat kecocokan antara informasi yang dicari oleh pengguna dengan hasil yang diberikan sistem. *Recall* adalah tingkat kesuksesan sistem dalam mendapatkan informasi kembali. Dan *accuracy* adalah tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya yang sudah diklasifikasi dengan benar. Perolehan nilai *accuracy* tertinggi sebesar 81% untuk tipe opini netral. Sedangkan nilai *precision* dan *recall* paling tinggi yaitu pada tipe opini negatif yaitu *precision* sebesar 53% dan *recall* sebesar 57%. Pada nilai *recall* dan *precision* memiliki hasil yang paling rendah untuk tipe opini netral karena data pada tipe opini netral hanya berjumlah 50. Berbeda dengan data pada tipe opini positif dan negatif yang berjumlah sama yaitu 150. Selain itu, banyak kalimat yang terklasifikasi sebagai tipe opini netral menjadi tergolong sebagai tipe opini positif atau negatif. Seperti contoh “harga rokok mahal aku tidak masalah, asal jangan harga paket internet yang makin mahal”. Kemudian pada contoh “dari pada menaikkan pajak rokok mending pajak mobil dan motor dinaikan, mengurangi macet polusi”. Dari hasil nilai *precision*, *recall*, dan *accuracy* pada tabel menunjukkan bahwa hasil pengujian belum dianggap baik. Hal ini bisa terjadi karena banyak data opini positif tergolong sebagai negatif maupun sebaliknya. Selain itu, juga banyak data opini yang terklasifikasi netral menjadi tergolong sebagai opini positif atau negatif. Berikut merupakan permasalahan yang menyebabkan hasil dari metode *Lexicon Based* belum dianggap baik.

### **3.1 Adanya Kata Ambigu**

Dalam melakukan penentuan, SentiWordNet mempunyai makna kata yang banyak dan nilai sentimen yang berbeda. Satu kata pada SentiWordNet dapat mewakili lebih dari satu makna sehingga dengan metode yang dipilih dapat menimbulkan kesalahan pemilihan makna kata. Karena pada *lexicon based* dengan metode yang dipilih yaitu *first sense* pencocokan frase dengan data sentimen yang ada pada dictionary menjadi bersifat tunggal. Artinya satu kata hanya memiliki satu sentimen yaitu pada kata yang muncul paling atas atau paling populer . Adanya kesalahan ini juga berakibat pada salahnya nilai sentimen. Seperti contoh pada *gold standart* data terdapat opini “harga rokok naik kemungkinan tindak kejahatan juga naik” sesudah diterjemahkan “cigarette prices go up possible crime also go up”. Pada kata “crime” memiliki dua makna kata dengan nilai sentimen netral dan negatif. Dengan metode First Sense, menimbulkan kesalahan klasifikasi karena dipilih nilai sentimen netral untuk kata “crime”. Selain itu pada contoh “setelah itu, harapan bagi orang yang tidak punya uang akan menghisap apa?” setelah diterjemahkan “after that, what to smoke for people with no money?”. Kata “no”

memiliki tiga makna pada nomor sense pertama yaitu bernilai “positif”, “netral” dan “negatif”. Namun nilai sentimen yang dipilih untuk kalimat tersebut adalah positif.

### **3.2 Penggunaan Kata-kata Tidak Baku**

Kata tidak baku merupakan kata yang dipakai tidak sesuai dengan kaidah bahasa yang sudah ditentukan. Kata tidak baku umumnya sering dijumpai pada percakapan sehari-hari. Banyak penggunaan kata-kata tidak baku dalam *gold standart* data yang tidak bisa terdeteksi oleh SentiWordNet. Kata “merokok” menjadi “ngerokok”, kata “uang” menjadi “duit”, kata “lebih baik” menjadi “mending”, dan masih banyak lagi. Selain itu juga ditemukan kata-kata tidak baku karena menggunakan bahasa daerah.

### **3.3 Fenomena Thwarted Expectation**

Fenomena *thwarted expectation* bisa terjadi pada keadaan dimana classifier mengalami kesulitan mendeteksi review (Pang & Lee, 2002). Seperti pada contoh “baguslah, saya suka tidak betah liat anak sekolah merokok, padahal dia masih kecil belum bisa nyari duit” pengguna menyampaikan opini positive untuk mendukung kenaikan harga rokok tetapi terdeteksi sebagai opini negative karena penggunaan kata-kata yang bernilai negative. Kemudian pada contoh “saya setuju jika harga rokok naik, bukan karena saya tidak merokok tetapi karena rokok di indonesia 80% untuk bergaya saja bukan untuk dinikmati”. Pengguna setuju dengan harga rokok naik dengan menuliskan kondisi realita yang jelek atas rokok namun juga terdeteksi sebagai opini negatif karena penggunaan kata-kata dengan nilai negatif. Banyaknya data yang terjadi karena fenomena ini turut mempengaruhi hasil klasifikasi yang kurang baik.

## **4. PENUTUP**

Penelitian kali ini merupakan analisis sentimen menggunakan metode *lexicon based* mengenai persepsi masyarakat terhadap kenaikan harga rokok pada media sosial twitter. Data opini ini disimpan sebagai *gold standart* data yang selanjutnya akan diklasifikasi menjadi 3 macam sentimen diantaranya positif, negatif, dan netral. Dari hasil penelitian, didapatkan perbandingan antara hasil sentimen yang dihasilkan oleh sistem dengan yang telah diklasifikasi secara manual masih belum dikatakan baik. Hal ini dikarenakan pada penelitian terjadi beberapa permasalahan. Perolehan nilai yang paling tinggi pada *precision* yaitu sebesar 53% dan *recall* sebesar 57% untuk tipe opini negatif dan perolehan nilai tertinggi *accuracy* sebesar 81% untuk tipe opini netral.

Untuk mendapatkan hasil klasifikasi terbaik, pengambilan data seharusnya berjumlah sama untuk tiap tipe opini. Seperti pada tipe opini netral yang hanya berjumlah 50 sebaiknya ditambah data lagi agar sesuai dengan jumlah data pada tipe opini positif dan tipe opini negatif

yaitu 150. Selain itu perlu dilakukan normalisasi kalimat pada permasalahan kata-kata tidak baku agar kalimat bisa dipakai sebagai bahasa yang sesuai dengan kaidah atau pedoman bahasa yang sudah ditentukan atau bersumber dari Kamus Besar Bahasa Indonesia (KBBI). Kemudian untuk adanya kata-kata ambigu dapat dilakukan dengan cara menentukan makna suatu kata atau *Word Sense Disambiguation* (WSD) sehingga makna kata polisemi atau kata yang memiliki makna lebih dari satu dapat teridentifikasi. Pada fenomena *thwarted expectation*, diperlukan suatu teknik untuk bisa mengidentifikasi kalimat sesuai topik tertentu.

## DAFTAR PUSTAKA

- Purba, I. D. C., Hidayati, H., & Gozali, A. A. (2014). *Metode Holistic Lexicon-Based untuk Analisis Sentimen pada Dokumen Bahasa Indonesia (Studi Kasus: Tweets mengenai Isu Sosial Kota Bandung)*. Tugas Akhir, Universitas Telkom.
- Buntoro, G. A., Adji, T. B., & Purnamasari, A. E. (2014). *Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation*. In The 6th Conference on Information Technology and Electrical Engineering (CITEE) (pp. 39-43).
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- Esuli, A., & Sebastiani, F. (2006, May). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of LREC (Vol. 6, pp. 417-422).
- Kumar, A., & Sebastian, T. M. (2012). *Sentiment analysis on twitter*. IJCSI International Journal of Computer Science Issues, 9(4), 372.
- Turney, P. D. (2002, July). Thumbs up or thumbs down?: *semantic orientation applied to unsupervised classification of reviews*. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. In LREC (Vol. 10, pp. 2200-2204).
- Pamungkas, E. W., & Putri, D. P. (2016). *An Experimental Study of Lexicon-Based Sentiment Analysis on Bahasa Indonesia*. Proceeding of The 6th International Annual Engineering Seminar (INAES).
- Listiana, Mila (2015) *Perbandingan Algoritma Decision Tree (C4.5) Dan Naïve Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus Puskesmas Kartasura)*. Skripsi thesis, Universitas Muhammadiyah Surakarta.

Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). *Thumbs up?: sentiment classification using machine learning techniques*. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.